DOCUMENT RESUME

ED 371 014 TM 021 645

AUTHOR Druva-Roush, Cynthia Ann; And Others
TITLE A Comparison of Decision Theoretic Approaches to Adjusting Cut Scores.
PUB DATE Apr 94
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994). Figures 1-5 are faint and will not reproduce well.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Comparative Analysis; *Cutting Scores; Decision Making; Discriminant Analysis; Grades (Scholastic); Higher Education; *Prediction; Psychometrics; Regression (Statistics); *Scoring; Statistical Studies; *Student Placement; *Validity
IDENTIFIERS *Contingency Tables; *Decision Theory

ABSTRACT

Methods of adjusting cut scores used in placement decisions are examined empirically. Admission and performance variables are used to study alternate methods of adjusting cut scores for placement in standard and accelerated rhetoric courses in a large university setting, with the predicted variable being success or failure as measured by end-of-semester course grades. Data from 4,045 freshmen entering a university in the fall semesters of 1991 and 1992 were collected (1,956 for setting cut scores, and 2,089 for validation). A comparison of predicted placement errors using cross validation groups is made between three decision theoretic approaches to setting cut scores: (1) contingency table analysis; (2) discriminant analysis; and (3) logistic regression. The effectiveness of each method is evaluated psychometrically, as is the ease of using each method in a practical setting. Two-equation discriminant analysis appears to be superior to logistic regression with more than one predictor variable, and it outperforms contingency table analysis. When a single index method is used, logistic regression and contingency table analysis are quite similar, and the contingency table analysis seems to be much more interpretable by nonstatisticians. (Contains 16 references, 6 tables, and 5 figures.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 371 014

A COMPARISON OF DECISION THEORETIC APPROACHES TO ADJUSTING CUT SCORES

by

Cynthia Ann Druva-Roush
Joyce E. Moore
Xiaoping Wang
Chi-Yu Huang
Ron Hueman

The University of Iowa
300 Jefferson Building
Iowa City, Iowa 52242-1470
(319)335-0356

American Educational Research Association
New Orleans
April 1994

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Cynthia Ann Druva-Roush

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 021645

BEST COP

Introduction

The purpose of this research is an empirical look at methods of adjusting cut scores used in placement decisions. Admission and performance variables are used to examine alternate methods of adjusting cut-scores for placement into standard and accelerated rhetoric courses in a large university setting. The predicted variable is success or failure as measured by end-of-semester course grades. A comparison of predicted placement errors using cross validation groups is made between three decision theoretic approaches to setting cut-scores: contingency table analysis, discriminant analysis, and logistic regression. The effectiveness of each method is evaluated pschometrically as well as the ease of using each method in a practical setting.

Related Research

Correct placement into standard and accelerated courses is a general concern for both faculty and students. Students within a given course should be relatively homogeneous in ability and encouraged to enroll in courses that are challenging but not beyond their ability. A number of methods for setting and adjusting cut scores have been presented in the competency testing literature. Willingham (1974) presents several models for defining competence and discusses them in detail. Berk (1986) provides a critical review of many of these methods. Meskauskas (1976) classifies these methods into either state or continuum models. State models assume that mastery/competence on true score performance is an all-or-nothing state (i.e. binary), while the continuum models assume that mastery/competency is a continuously distributed ability that can be viewed as an interval

Y AVAILABLE

circumscribing the boundaries for mastery/competence (i.e. continuous). Hambleton and Eignor (1980) further subdivide continuum models into three categories labeled judgmental, judgmental-empirical, and empirical-judgmental. Judgmental methods are based entirely on the judgments of one or more persons. Judgmental-empirical methods are based primarily on the judgments of one or more persons with performance data made available to guide those judgments. Empirical-judgmental methods are based primarily on performance data from one or more groups of examinees and the statistical analysis of those data.

Of the empirical-judgmental methods less than half of those reviewed by Berk deal with adjusting standards as opposed to setting standards. Jaeger (1989) provides an overview of standard setting procedures based on experimental-judgmental models. The empirical-judgmental methods for adjusting standards does not determine a standard; rather, they presume that a standard already exists on an external criterion and merely translates this into a cut-off score on the test (Shephard, 1980). These methods for locating the passing score on the test are based on decision theory. The object is to match the test dichotomy to the criterion dichotomy to ensure the smallest number of classification errors.

Comparison in outcomes of the judgmental methods have been made by Glass (1978), Hambleton (1978), Skakun and Kling (1980), Berk (1986), Mills and Melican (1988). Of the empirical-judgmental methods, Berk (1986) examines only the standard setting approaches. Little research has been performed comparing empirical judgmental statistical decision models for adjusting cut-scores in an educational setting

Methods

This study will examine the placement accuracy of three methods: 1) contingency table analysis, 2) discriminant analysis, and 3) logistic regression. The contingency table analysis has as its theoretical basis the work of Berk (1976) on setting cut-scores for criterion-referenced measures. For this procedure, "success" and "failure" performance levels on the criterion variable (e.g. course grade) are set. For a given score, the percentage of students that are incorrectly classified (either false negatives - predicted to fail but instead succeeded; or false positives - predicted to succeed but instead failed) are added together. The cut-score is set where the sum of these errors is minimized. This approach offers easy interpretation, but is limited to use with only one predictor variable. Regression methods that allow for more than one variable to predict binary outcomes include discriminant analysis and logistic regression.

With discriminant analysis, the object is to classify a subject into one of two groups.

> "Fisher's (1936) idea was to transform the multivariate problem into a univariate one, in the sense of finding the linear combination of the x's (a single composite variable) which will maximally discriminate the groups. ... It is assumed that the two populations are multivariate normal and have the same covariance matrix." (Stevens, 1992, p. 289)

The intent is to include only those predictor variables with their appropriate weights that maximally discriminate between groups. Two sets of classification function coefficients (Fisher's linear discriminant functions) are calculated using as a condition to enter

a statistically significant Wilk's lambda statistic. For each case, the classification scores for each group are computed from the classification function coefficients. The case is then classified into that group in which the classification score is highest. An approximation to this method involves use of standardized canonical discriminant function coefficients. The average values of the predictor variables for the two different groups are placed into a single function which results in two different values. The cut is then set at the midpoint of these two values (Stevens, 1992)

In logistic regression, an attempt is made to estimate the probability of a case being in a particular group through use of a log-linear model. For example, if Y=1 represents a student who is successful in a particular course and $\mathbf{x}$ is the vector of predictor variables under study, $P[Y=1|\mathbf{x}] = e^{g(\mathbf{x})} / \{1 + e^{g(\mathbf{x})}\}$ where $g(\mathbf{x}) = \beta_0 + \sum \beta_i x_i$. An index is created from a weighted combination of relevant variables with the cut score set at that index value where the proportion of correct placement reaches 50%. Efron (1975) and Press and Wilson (1978) suggest that unless the data depart from the assumptions required by the linear regression model (normal distributions and identical covariances), discriminant analysis is to be preferred over logistic regression.

These three different decision theoretic approaches to adjusting cut scores were applied to the data. The percent of correct placement in a cross validation group, the nature of incorrect placement (percentage of false positives vs. false negatives) as well as the cut scores themselves were used for comparison between methods.

Data Source

Data from 4045 freshmen entering the University during the 1991 and 1992 fall semesters who took either the standard rhetoric course involving two semesters or the one semester accelerated course was collected (n=1956 for setting cut-score and n=2089 for validation). A randomly selected validation group was used to follow-up on the success and nature of errors for each placement method. Placement prior to 1990 was based on a weighted combination of the "old" ACT English and Social Studies sub scores. The 1991 fall semester provided the first opportunity to examine the relationship between enhanced ACT scores and success and failure in Rhetoric courses. Success is defined as a B- or above as a course grade. Predictor variables included in the initial research were enhanced ACT sub scores and composite, and high school rank. High school GPA was not included due to a high percentage of missing data.

Results

Descriptive Statistics

The descriptive statistics for the two variables which most closely correlated with course grades (ACT English (ACT-E) and High School Rank (Rank) are displayed in Table 1. The data for 1991 and 1992 are quite different. The average ACT English subscores and high school percentile ranks were much lower for 1992 Rhetoric classes. In addition, there was a much larger spread of ACT-E subscores for 1992. Is it feasible to assume that one method for adjusting out scores is equally precise when data may vary significantly from year to year?

Comparison of Covariance Matrices

The results of comparing covariance matrices of those that succeeded in a class (passes) with those that did not succeed (fails) are displayed in Table 2 by year and by course. These results indicate that for both Rhetoric 10:1 and 10:3, there was a significant difference in the pass and fail covariance matrices for 1991 data but not 1992.

More than One Variable

Logistic regression and discriminant analysis were each used as methods to combine multivariate predictors to predict success or failure in either Rhetoric 10:1 or Rhetoric 10:3. Each course was examined separately. Although all ACT scores and high school rank were available, only the ACT English (ACT-E) and high school percentile rank (Rank) entered into the final equations for Rhetoric 10:1. For Rhetoric 10:3, ACT-E did not enter, with high school rank being the only viable predictor in the discriminant analysis. (ACT math did enter but did not seem to represent to the English department a logical predictor.) Figures 1 and 2 show the results of the discriminant analysis and logistic regressions using 1992 data. Tables 3 and 4 display the cuts set by these two methods, hit rates, and types of error made on the validation groups for each course and year.

Discriminant Analysis

As was mentioned earlier in the methods section, there are two methods of decision making available with discriminant analysis: 1) creating two equations for classification or 2) setting a single cut score. The SPSS-X discriminant analysis program automatically sorts the data into an analysis group and a validation group. A followup is then performed using the first method of classification. The second method involves calculating the cut

score and then applying it to the pre-selected validation group. The results from the second method were then compared with the results from the logistic regression. In reference to Tables 3 and 4, the two equation method of classification is, in every case, more accurate than the approximation of a single index. The simplification of two equations to a single index creates a much higher proportion of false negatives (incorrectly failure when in fact a student succeeds).

With reference to the assumptions required with discriminant analysis, 1991 has covariance matrices that are statistically significantly dissimilar, particularly in Rhetoric 10:3. There is a significant decrease in hit rate in both methods from 1992 to 1991 as would be predicted (Efron, 1975; Press and Wilson, 1978).

Logistic Regression

Figure 2 displays the probability curves for the 1992 Rhetoric data using logistic regression. The cut is set at the index value at which a case has a 50% probability of being in the pass group. Table 3 indicates that for Rhetoric 10:1, the hit rates are slightly higher for the discriminant analysis two equation procedure than the logistic regression in both 1991 and 1992. Even when the covariance matrices were significantly different (i.e. 1991), the discriminant analysis outperformed the logistic regression. For Rhetoric 10:3, displayed in Table 4, only Rank entered into both the logistic regression and discriminant analysis for 1992. For 1991, both predictive variables, ACT-E and Rank, entered into the discriminant analysis. The logistic regression cut scores for Rhetoric 10:3, based only on rank, seem to be extremely low. However, when the discriminant analysis is reduced to a single cut score, the hit rate, in comparison, with the logistic regression, drops off dramatically.

One Variable

Although Rank had a higher Pearson Product Moment Correlation Coefficient, ACT-E seemed to the English department to represent the single variable to be used for placement into the entry level Rhetoric courses. Figures 3, 4 and 5 display the results of the analysis of 1992 data for the three methods of analysis: logistic regression, discriminant analysis, and contingency table analysis. Table 5 shows the cuts and hit and error rates for Rhetoric 10:1 data. As was seen earlier, the two equation method of discriminant analysis has a higher percentage of correct placements than the approximated single index cut. The two equation method of discriminant analysis using a single variable outperforms the logistic regression and contingency table. In 1992, the logistic regression slightly outperforms the contingency table analysis with a slightly higher successful placement rate as well as a lower false negative error rate. Logistic regression and contingency table analysis have identical cuts for the 1991 data.

Table 6 shows the cuts and hit and error rates for Rhetoric 10:3 data. Because of the similarity in ACT-E between the group which succeeded (receiving a B- or above in Rhetoric 10:3) and the group which failed (less than a B-) in 1992, a discriminant analysis using a single variable could not be performed. The logistic regression and contingency table analysis derived significantly different cut scores but similar hit rates with logistic regression having a slightly higher false positive error rate and a lower false negative error rate. In 1991, hit rates were quite similar between the discriminant analysis two equation method, logistic regression, and contingency table analysis. The discriminant analysis one index cut method is inferior to the other two methods of decision making.

Discussion

Placement tests are administered to a substantial number of incoming freshmen each semester in areas such as foreign language, mathematics or chemistry. Placement test results and admissions information are routinely used to place students into appropriate courses their first semester. The efficiency and accuracy of this process depend on a reliable method for setting cut scores. Previous work has suggest that logistic regression is the appropriate method to use for placement purposes in an educational setting (Crouse, 1991). This paper reports the result of an empirical search for the best method for placing students into either a two semester Rhetoric course (10:1) or an accelerated one semester Rhetoric course (10:3). Because some a pre-existing selection method was already in place (based upon old ACT scores and their concordance with the new ACT scores), some restriction in range was present in both 1991 and 1992. For each course, there is a greater number of successes to failures, so the distributions are disproportionate and in some cases skewed.

In this study, an examination of methods involving more than one predictor variable, shows that the two equation discriminant analysis is superior to the logistic regression method. This finding held even when the covariance matrices were not statistically similar. However, discriminant analysis would require an academic advisor to calculate two values for each individual for placement. A more manageable scenario for advisors would be a single equation or cut score. For practical reasons logistic regression is preferable to relying on discriminant analysis.

Although the use of more than one predictor variable resulted in slightly higher correct placement, use of a single content-similar measurement is more interpretable for the user. In this case, the two equation discriminant analysis method

outperformed both logistic regression and contingency table analysis. Again, when a single index method is used, logistic regression and contingency table analysis are quite similar. In addition, the contingency table analysis process seems to be much more interpretable by non-statisticians than does logistic regression.

References

Berk, R.A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. Review of Educational Research, 56, 237-172.

Crouse, J. D.(1992) The Validity of ACT Assessment Scores for Freshmen Course Placement at the University of Iowa. Paper prepared by ACT for The University of Iowa.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. Journal of the American Statistical Association, 70, 892-898.

Fisher, R.A. (1936). The use of multiple measurement in taxonomic problems. Annals of Eugenics, 7, 179-188.

Hambleton, R.K. (1978). On the use of cut-off scores with criterion-referenced tests in instructional settings. Journal of Educational Measurement, 15, 277-290.

Hambleton, R.K. and Eignor, D.R. (1980). Competency test development, validation, and standard setting. In R.M. Jaeger and C.K. Tittle (Eds.), Minimum competency achievement testing: Motives, models, measures, and consequences. Berkeley, CA: McCutchan.

Jaeger, R.M. (1989). Certification of student competence. In R.L. Linn (Ed.) Educational Measurement, 3rd edition, Washington, D.C.: The American Council on Education, 485-514.

Meskauskas, J.A. (1976). Evaluation models for criterion-referenced testing: Views regarding mastery and standard setting. Review of Educational Research, 45, 133-158.

Mills, C.N. and Melican, G.J. (1988). Estimating and adjusting cutoff scores: Features of selected methods. Applied Measurement in Education, 1, 261-275.

Press, J. and Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association, 73, 699-705.

Shepard, L.A. (1980) Technical issues in minimum competency testing. In D.C. Berliner (Ed.), Review of research in education (vol.8). Washington, D.C.: American Educational Research Association, Shepard, L.A. , 30-82.

Shepard, L.A. (1983) Standards for placement and certification. In S.B. Anderson and J.S. Helmick (Eds.), On educational testing. San Francisco: Jossey-Bass, 61-90.

Shepard, L.A. (1984). Setting performance standards. In R.A. Berk (Ed.), A guide to criterion-referenced test construction. Baltimore, MD: The Johns Hopkins University Press, 169-198.

Skakun, E.N. and Kling, S. (1980). Comparability of methods for setting standards. Journal of Educational Measurement, 17, 229-235.

Stevens, J. (1992). Applied multivariate statistics for the social sciences: second edition. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Willingham, W.W. (1974). College Placement and Exemption. New York: College Entrance Examination Board.

Table 1

Descriptive Statistics for 1991, 1992, and 1993 ACT English Scores and High School Rank for Rhetoric 10:1 and Rhetoric 10:3 by Pass (P) and Fail (F) Classification

	Variable					
		ACT English		Rank		
	n	Average	Standard Deviation	Average	Standard Deviation	Correlation
Year						
1991						
10:1 F	160	20.27	2.73	57.74	16.14	.095
P	456	21.48	2.29	69.63	15.92	
10:3 F	62	26.45	2.22	67.55	18.83	.136
P	189	27.84	2.61	83.02	14.27	
1992						
10:1 F	257	18.50	6.53	63.78	15.50	.048
P	523	19.65	6.07	71.94	15.85	
10:3 F	78	25.62	4.85	71.19	17.10	.123
P	347	25.96	6.32	81.85	16.33	

Table 2

Test of Equality of Group Covariance Matrices Using Box's M

		Box's M	F	df	Significance
Year	Course				
1991					
	10:1	7.8413	2.6008	3	.0506
	10:3	10.551	3.4719	3	.0155
1992					
	10:1	3.5608	1.1831	3	.3147
	10:3	.27148	.27026	3	.6032

Table 3

Results of Discriminant Analysis and Logistic Regression Performed for 1991 and 1992 for Rhetoric 10:1

		Method of Analysis		
		Discriminant Analysis		Logistic Regression
		Two Equations	Cut Score	
Year				
1991				
	n	772	780	781
	Cut		69.52	
	Hits	67.88%	58.26%	63.51%
	False +	26.81	10.63	24.71
	False -	5.31	31.11	11.78
1992				
	n	606	616	615
	Cut		62.49	
	Hits	76.57%	61.75%	73.91%
	False +	20.63	11.26	25.43
	False -	2.80	26.99	.66

8

Table 4

Results of Discriminant Analysis and Logistic Regression Performed for 1991 and 1992 for Rhetoric 10:3

		Method of Analysis		
		Discriminant Analysis		Logistic Regression
		Two Equations	Cut Score	
Year				
1991				
	n	393	425	427
	Cut	Rank Only	79.90	35 (Rank Only)
	Hits	80.15%	64.54%	76.59%
	False +	18.58	8.41	18.64
	False -	1.27	27.05	4.09
1992				
	n	251	264	264
	Cut		76.02	45 (Rank Only)
	Hits	75.70%	68.56%	76.52%
	False +	15.94	9.09	20.45
	False -	8.37	22.35	3.03

Table 5

Results of Discriminant Analysis, Logistic Regression and Contingency Table Analysis Performed for 1991 and 1992 for Rhetoric 10:1 Using ACT-E As a Predictive Variable

Method of Analysis		Discriminant Analysis		Logistic Regression	Contingency Table Analysis
		Two Equations	Cut Score		
Year					
1991					
	n	589	606	604	604
	Cut		20.88	19	19
	Hits	77.06%	63.91%	71.85%	71.85%
	False +	22.44	14.07	19.54	19.54
	False -	5.00	22.02	8.61	8.61
1992					
	n	772	737	781	781
	Cut		19.08	16	17
	Hits	69.30%	59.15%	65.17%	63.38%
	False +	30.70	19.85	27.91	27.53
	False -	0	21.00	6.91	9.09

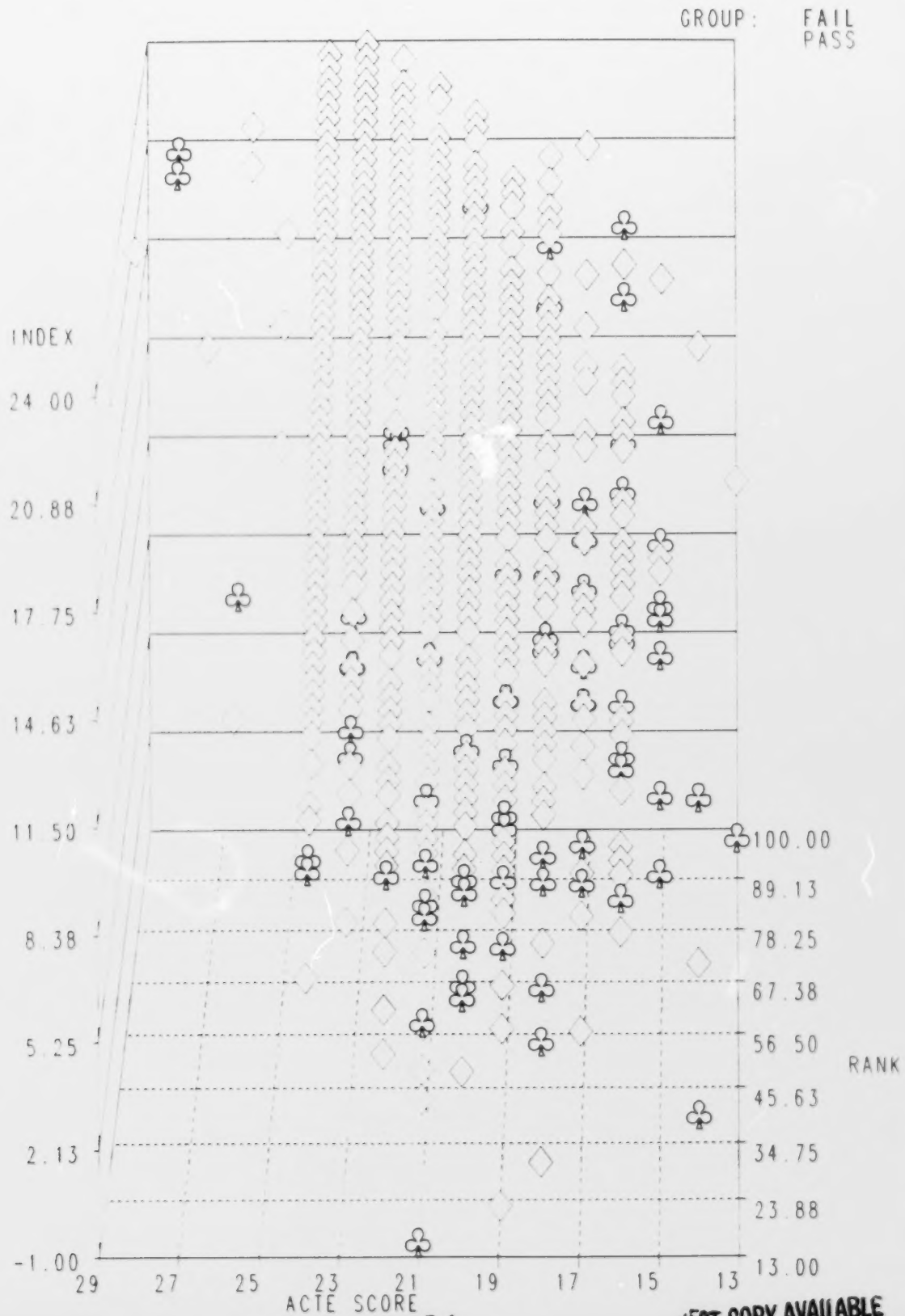
Table 6

Results of Discriminant Analysis, Logistic Regression and Contingency Table Analysis Performed for 1991 and 1992 for Rhetoric 10:3 Using ACT-E As a Predictive Variable

		Method of Analysis			
		Discriminant Analysis		Logistic Regression	Contigency Table Analysis
		Two Equations	Cut Score		
Year					
1991					
	n	261	251	264	264
	Cut		27.14	25	24
	Hits	75.70%	53.79%	75.38%	75.00%
	False +	22.31	5.68	21.59	22.73
	False -	1.99	40.53	3.03	2.27
1992					
	n	No Variables Qualified		440	440
	Cut			18	23
	Hits			77.27%	76.14%
	False +			18.64	17.95
	False -			4.09	5.91

Figure 1

Discriminant analysis (1992) for Rhetoric 10:1 using both ACT-E and Rank.



BEST COPY AVAILABLE

Figure 2

Logistic regression (1992) for Rhetoric 10:1 using both ACT-E and Rank.

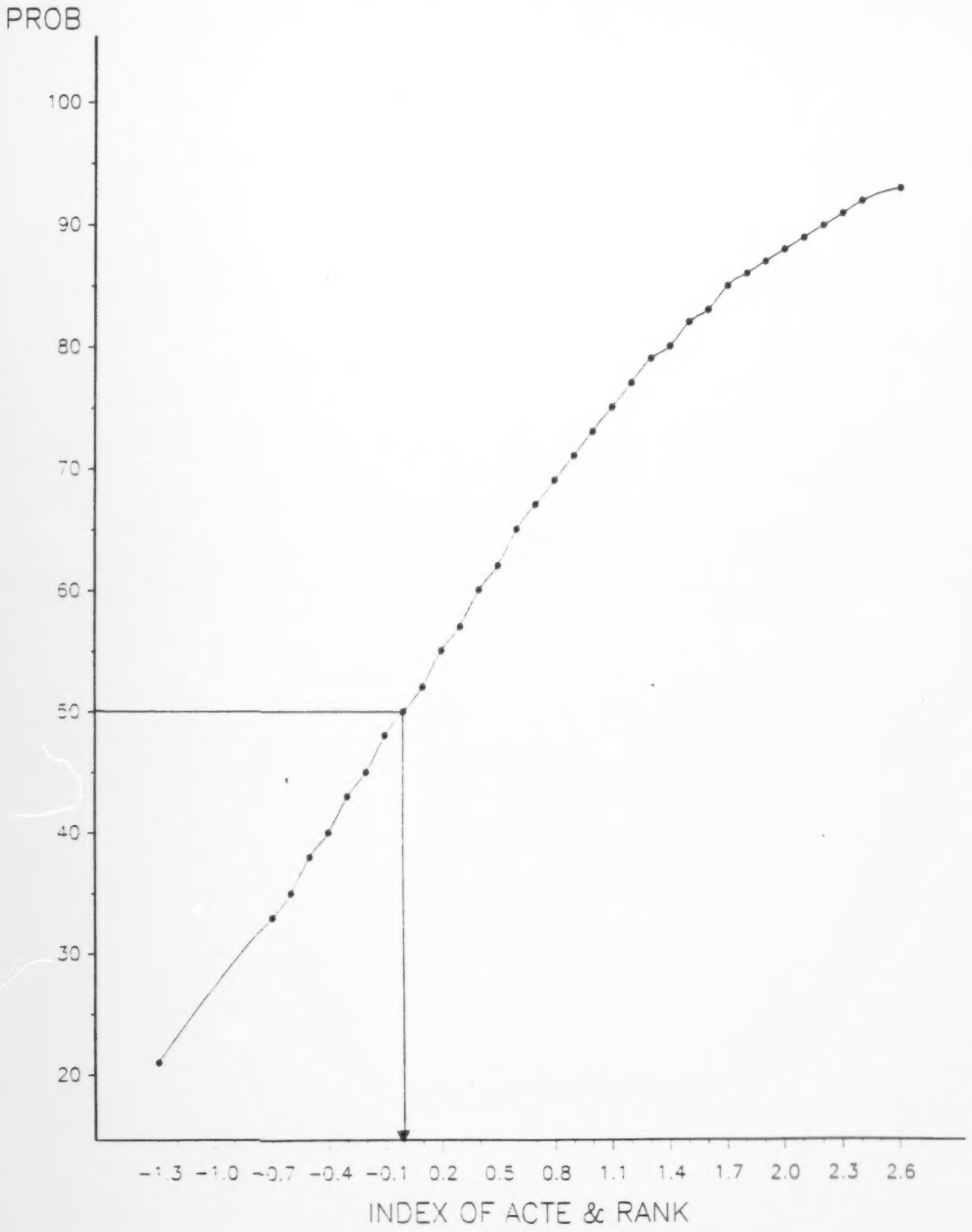
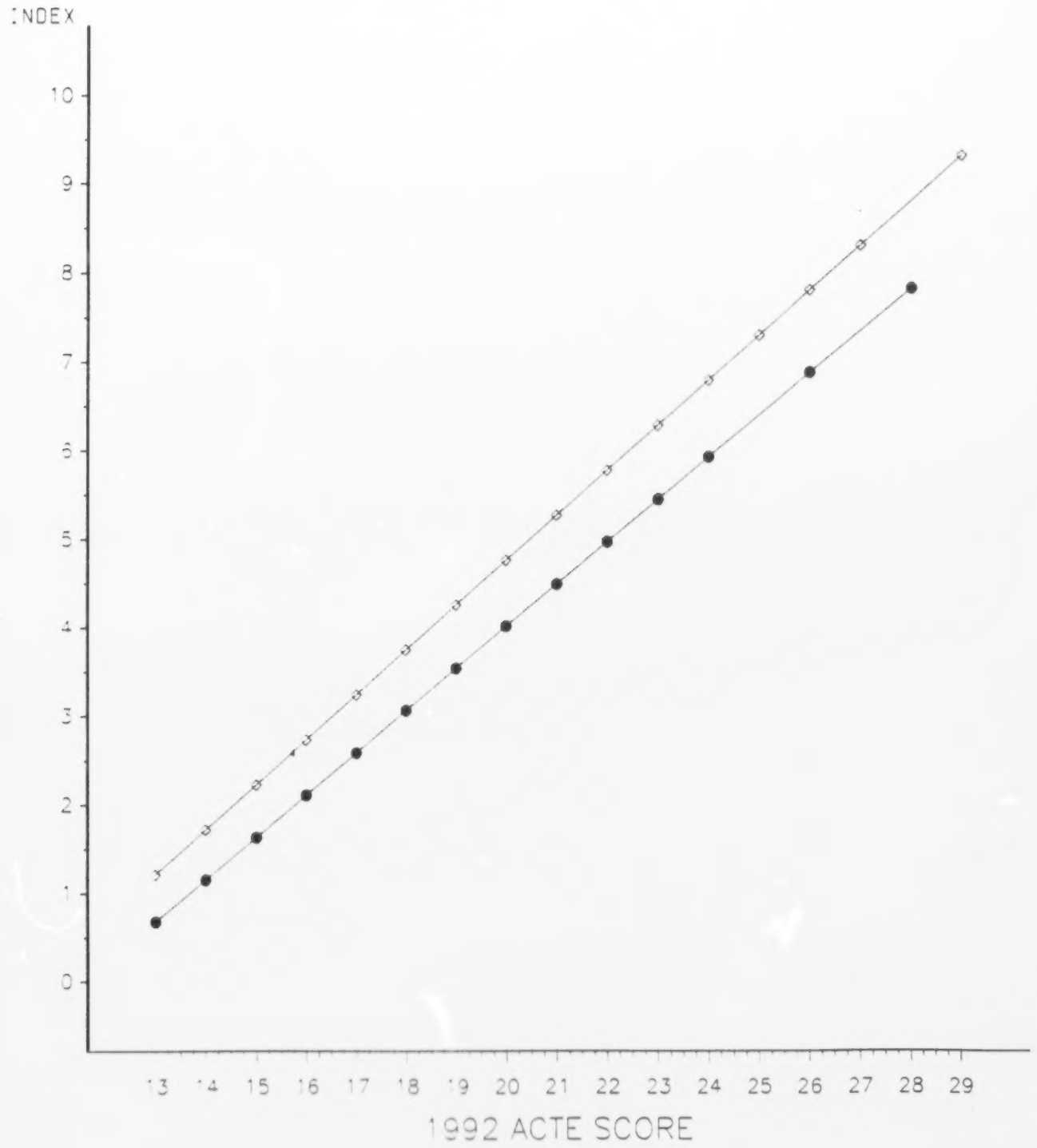


Figure 3

Discriminant anlysis (1992) for Rhetoric 10:1 using ACT-E as predictor



● ● ● FAIL ◇ ◇ ◇ PASS

Figure 4

Logistic regression (1992) for Rhetoric 10:1 and 10:3 using ACT-E as

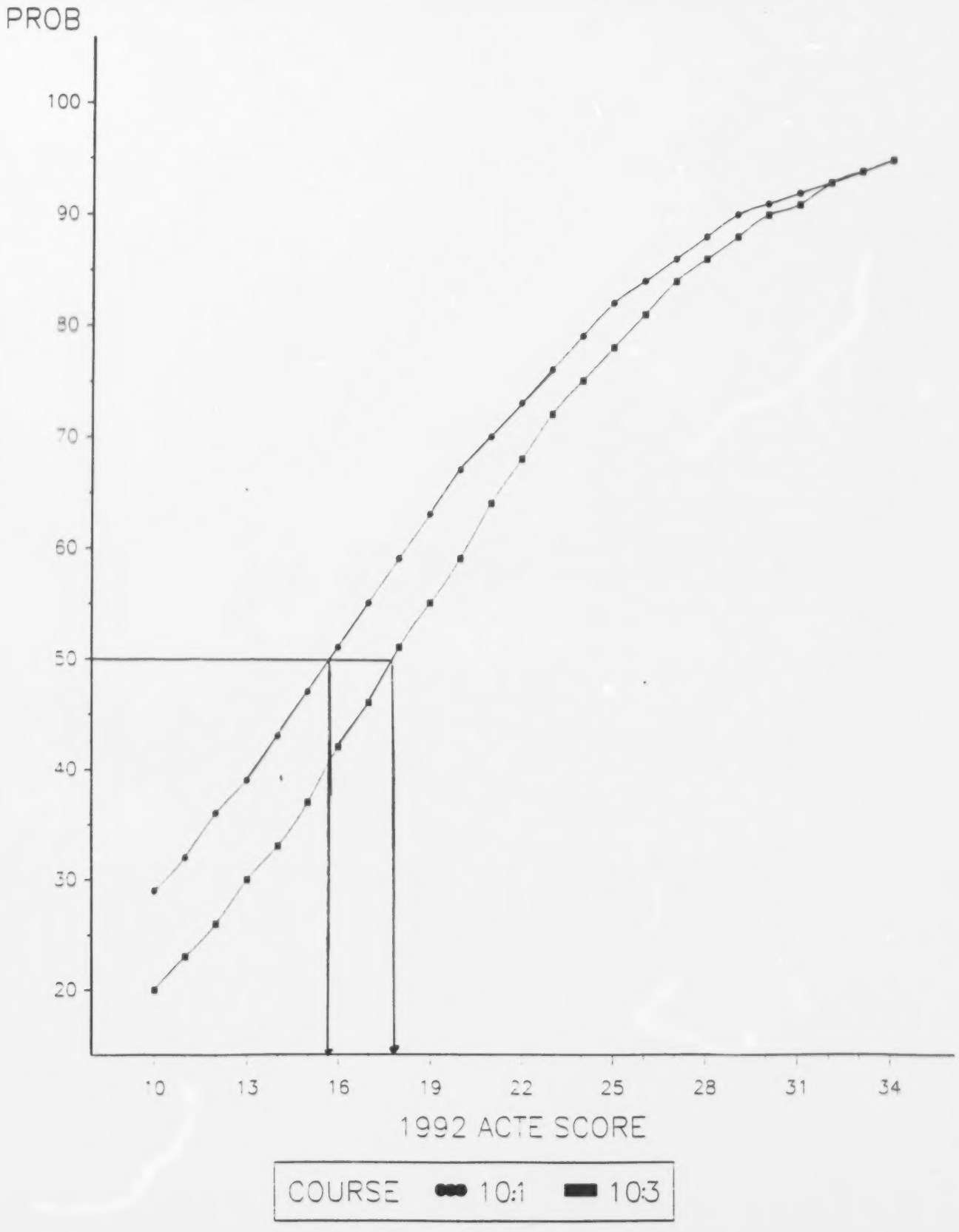


Figure 5

Contingency table analysis (1992) for Rhetoric 10:1 and 10:3 using ACT-